Linked open data for agricultural land evaluation

Charitas Fibriani[1*], *Mardhani* R.[1], and *Ahmad* Ashari[1]

[1]Universitas Gadjah Mada, Sleman, Yogyakarta, Indonesia

Abstract. The government has programs that related to food independence which can be achieved by optimizing agricultural data analysis as decision support. The analysis requires a combination of several data but these data are in different sources and formats. Linked Open Data (LOD) is the answer, it makes the interconnectedness of all data that be needed, the interconnected data can be used for the next process in land evaluation. Land evaluation has 2 activities: land characteristics, and land quality. This research focus on designing a prototype of a land evaluation system model based on linked open data. There are 12 data sources that will be used for this land evaluation. The datas are spatial data which have longitude and latitude. In addition, any 2 types of data that be used text and geojson. The linked between these attributes from many sources and was using LOD's consept creates a relationship terminology.

1 Introduction

The government has programs that related to food independence which can be achieved by optimizing agricultural data analysis as decision support. Agricultural products must be developed in areas that are in accordance with their characteristics and have suitable land quality for them so that agricultural production results get maximum results.

The analysis was carried out to find alternative solutions to increase agricultural production yields [1]. The search for alternative solutions requires a database subsystem, model base subsystem and interactions between these subsystems [2]. Land evaluation is a solution to get alternative solutions to get maximum agricultural yields. Land evaluation is obtained from 3 factors, namely: (1) Land Use Type; (2) Land Characteristics, (3) Land Quality.

The data used in this study is spatial data related to land evaluation, namely land characteristics and land quality factors. Where the spatial data related to land evaluation is found in various sources with various formats. The thing that needs to be considered in finding information using data on the internet is that the data needed are in different sources or different databases [3] with different business processes, data structures and distributed system designs. Distributed system design is the architecture of the system that connects these data. The spatial data related to land evaluation are not related to each other, because the data sources are different from one another, the characteristics and variations of the data types are different [4].

* Corresponding author: charitas.fibriani@mail.ugm.ac.id

Agriculture requires analysis by utilizing a combination of several data [5], but these datas are in different sources and formats. The Berners Lee team has presented the development of a five-star rating system on open data technology[6]. This allows the data related to agriculture that is needed is already available on the web and is open data. In fact, if the spatial data on the open data portal are interconnected, it can be used to answer an agricultural problem from the soil testing process to the harvesting process. Linked Open Data (LOD) technology is the answer to the problem, because LOD creates a relationship between data from various sources and different type.

The principle of Linked Data can be the bridge of the spatial data related for evaluation of the land. Where differences in data sources and differences in formats are no longer a problem. So that with the interconnectedness of these data, it can be used for the next process in land evaluation.

2 Literature review

There are several studies that discuss the linked design between datasets, the examples are Ashari [7], Deng [8] and Huuskonen and Oksanen [9]. However in this researches, the design between data sets that contain the relationship between them is not discussed.

Meanwhile research Cifuentes-silva et al. [10] and Roussey et al. [11] have discussed to design architectural models for LOD. The standard level of data used in this study is at the standard level (*) to level (***) and the data types used are the same.

Research that has used the standard level (****) is Becker and Christian [12] and Kozák et al. [13]. The result of Becker and Christian [12] is an application that provides information based on promiliarity or proximity to the user's location. This information is obtained by utilizing the geographical coordinates of the user and then connecting with data that has been stored from various sources and has been connected to each other using datasets from DBpedia.

Meanwhile, the research Kozák et al. [13] and resulted a linked open data design that connected several sources related to drug names, indications and contraindications of a drug. The concept used is Natural Language Programming by detecting the similarity of names.

The development of open data technology makes data published openly on the internet [14], but to obtain a relevant information it need more effort. Five-Stars Level System is a classification for data structures, which are divided into 5 levels [15]. At Level (*) the data is available on the web in any format, an example of the data is a scanned image on the web. Level (**), users at this level can directly process data with the help of certain software (calculating processes, displaying charts, and so on). The disadvantage is that data at this level is still difficult to read because it requires certain software to open it. Data at this level is also easy for uploaders to upload. Level (***), data is available in non-proprietary format. Users at this level can easily retrieve data from the web and can manipulate the data without relying on specific software. Level (****), the data has been uploaded using the open data standard from the World Wide Web Corcortium. The data item has a Uniform Resource Identifier (URI) that can be shared on the web. However, because it uses the Resource Decryption Framework (RDF) to present data, users at this level must understand the RDF structure. Data at Level (*****) meets the requirements from data at Level (*) to data at Level (****) plus that data at this level has links with other data. In Table 1, it can be seen that the five-star level system represents the level of open data at each level.

Table 1. Five-Stars level system [6].

Stars	Description
*	Open data with open license, in any data type (example data: *.jpg)
**	Data in machine readable form (example: *.xlx)
***	Non-proprietary data format (example: *.csv)
****	Same with the three levels above but the data must have RDF.
*****	Same as the four levels above but the data are interrelated

3 Land evaluation procedure

Land evaluation is a follow-up activity, which in this paper will focus on 2 activities that must have existed before, these activities are: (1) Land Characteristics, and (2) Land Quality. These activities are obtained from the results of measurements in the field by various parties. So that the data is stored in various sources with different formats.

Land characteristics are properties of a land which are characteristics attached to the land. The characteristics of the land used can be seen in Table 2.

Table 2. Attributes of land characteristics.

No	Atributte	Description
1	Annual average temperature	Annual average temperature Annual average air temperature
2	Rainfall	Precipitation Amount of annual precipitation
3	Air Humidity	Air Humidity Air humidity level
4	Drainage	Drainage Effect of water percolation rate into the soil on air aeration in the soil
5	Texture	Texture Comparison of sand grains
6	Rough Material	Coarse Materials >2 mm
7	Effective Depth	Effective Depth The depth of the soil layer that can be utilized for the development of plant roots
8	Peat Maturity	Peat Maturity Fiber content level
9	Peat Thickness	Peat Thickness Level of peat thickness
10	Land CEC	Soil CEC Ability of soil to exchange cations
11	Base Saturation (KB)	Base Saturation (KB) Total base base extracted from NK_4OAC in every 100 grams of soil sample
12	Soil pH	Soil pH is the [H+] in the soil solution
13	C-organic	C-organic Organic carbon content in the soil
14	Total N	Total N Total N content in the soil
15	P_2O_5	P_2O_5 Content of 25% HCI extracted K_2O_5 in soil
16	K_2O	K_2O Content of 25% HCI extracted K_2O in soil
17	Salinity	Salinity The amount of salt content is easily soluble in the soil
18	Alkalinity	Alkalinity The amount of sodium (Na) content
19	Specific Depth	Specific Depth The depth of the sulfidic material is measured from the soil surface to the upper boundary of the sulfidic layer
20	Slope	Land Slope
21	Rocks on the Surface	Rock at Surface Volume of rock at ground level
22	Rock Outcrop	Rock Outcrop Volume of rock that appears on the ground
23	Landslide Danger	Landslide Hazard Movement of rock or soil mass
24	Erosion Danger	Erosion Hazard Amount of soil lost from a field
25	puddle	Inundation State the height and duration of inundation.

Land quality is an identifying property of a land. Land quality has several attributes in common with land characteristics. The attributes on land quality can be seen in Table 3.

Table 3. Land quality attributes.

No	Atributte	Description
1	temperature (tc)	Determined by the annual average air temperature
2	Water Availability (wa)	Determined by rainfall, humidity and agro-climatic zone
3	Oxygen Availability (oa)	Determined by drainage
4	Rooting Media (rc)	Determined by drainage, texture, coarse material, effective depth, maturity and thickness of peat
5	Nutrient Retention (nr)	Determined by soil CEC, KB, pH and organic C
6	Nutri available (na)	Determined by total N, P_2O_5 and K_2O
7	Toxicity (xc)	Determined by salinity
8	Sodicity (xn)	Determined by Alkanility
9	Sulfidic Hazard (xs)	Determined by sulphidic depth
10	Erosion Hazard Level (uh)	Determined by erosion hazard and soil depth
11	Danger of Flood/Inundation (fh)	Determined by the height and length of inundation
12	Land Preparation (lp)	Defined by rocks on the surface and rock outcrops.

There are 4 land suitability classifications, namely: (1) Order, (2) Class, (3) Subclass and (4) Unit. The order describes land in general and simple, there are only 2 values, namely “Suitable” (S) or “Not Suitable” (N). The class describes the level of land suitability within the order, the value in the order “Suitable” is divided into “Very Suitable” (S1), “Sufficiently Appropriate” (S2), and “Marginal Appropriate” (S3). While the value in the order "Not Appropriate" is not distinguished. The subclass describes the level of land suitability within the class, where the maximum limiting factor is only two so that it is still possible to improve the land. While the Unit describes the level of land suitability in the subclass. The unit is the distinguishing factor from the limiting factor.

4 Linked open data for land evaluation

This system is a prototype of a land evaluation system model based on linked open data. There are 12 data sources that will be used for this land evaluation. Where there are 2 types of spatial data used, text and geojson. Data on source A1 has geojson data type, while data on source A2, A3, A4, A5, A6, A7, A8, A9, A10, A11, A12 and A13 has text format.

Source A1 contains regional spatial data in geojson, namely polygons containing administrative information for the coordinate area. While sources A2-A13 contain spatial data in text. A14 is a display to get the results of the linked data from the sources A1 to A13. For more details regarding the information contained in each source, see Table 4.

Table 4. Sources of land evaluation data.

Source	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13
Longitude	√	√	√	√	√	√	√	√	√	√	√	√	√
Latitude	√	√	√	√	√	√	√	√	√	√	√	√	√
Ward	√												
districts	√												
district	√												
temperature		√											
Rainfall			√										
Humidity			√										
Drainage				√	√								
texture					√								
Rough material					√								
Effective depth					√						√	√	
Peat maturity					√								
Peat thickness					√								
Land CEC						√							
Language saturation						√							
Soil pH						√							
C-organic						√							
Total N							√						
P2O5							√						
K2O							√						
Salinity								√					
Alkalinity									√				
Specific depth										√			
Slope											√		
Rocks on the surface													√
Rock Outcrop													√
Landslide hazard											√		
Erosion hazard											√		
puddle												√	

The relationship between these attributes creates a relationship terminology. Where each data source relates to other data sources using a relationship terminology, for example source A2 has a "get in" id relationship on source A1. A14 gets the "id" on sources A2 to A13 then relate to source A1, by getting the id, we get the spatial location of the region. The relationship between sources A1, A2, A3, A4, A5, A6, A7, A8, A9, A10, A11 and A12, A13 and A14 can be seen in Figure 1.

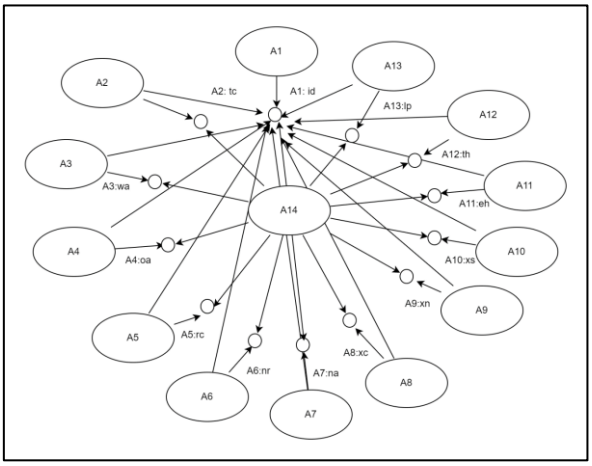

Fig. 1. Data set architecture and linked data for agricultural evaluation.

5 Conclusion

Linked Open Data for land evaluation can use the factors of land use type, land characteristics and land quality. Where the three factors have attributes of longitude, latitude, kelurahan, sub-district, city/district, land use data, yield, market orientation, capital intensity, labor intensity, land ownership status, income level, land use, yield, market orientation, intensity capital, labor intensity, land ownership status, income level, land management, technical knowledge, farmer culture, land management technology, infrastructure needs, farm area, temperature, water availability, oxygen availability, root media, nutrient retention, available nutrients , toxicity, sodicity, sulfidic hazard, erosion hazard, flood/inundation hazard and land preparation.

The longitude, latitude, kelurahan, sub-district, and district are attributes that are used as bridges to connect sources related to the evaluation of the land. Where the relationship is created because of the terminology of the relationship on the attributes in the source.

References

1. A. Nada, IEEE (2014)
2. G. Rui and L. Yongyi, 55 (2016)
3. S. Chawathe, H. Garcia-molina, J. Hammer, K. Ireland, Y. Papakonstantinou, J. Ullman, and J. Widom, IPSJ 1 (1994)
4. L. Bouganim, T. Chan-sine-ying, J. Darroux, G. Gardarin, and F. Sha, in *Proc. 25th VLDB Conf. Edinburgh. Scotl.* (1999), pp. 750–753
5. A. Alrefae and J. Cao, in *Proc. IEEE Int. Conf. Comput. Netw. Informatics, ICCNI 2017* (2017), pp. 1–10

6. H. Linden and J. Strale, KTH (Royal Inst. Technol. Sch. Technol. Heal. 136 40 Handen, Sweden (2015)
7. A. Ashari, **7**, 449 (2016)
8. D. Deng, *Construction and Reuse of Linked Agriculture Data : An Experience of Taiwan Government Open Data* (Springer International Publishing, 2018)
9. J. Huuskonen and T. Oksanen, Comput. Electron. Agric. **154**, 25 (2018)
10. F. Cifuentes-silva, C. Sifaqui, B. Congreso, and J. E. Labra-gayo, 79 (2011)
11. C. Roussey, S. Bernard, F. Pinet, X. Reboud, V. Cellier, I. Sivadon, D. Simonneau, and A. Bourigault, Comput. Electron. Agric. **142**, 632 (2017)
12. C. Becker and B. Christian, **7**, 278 (2009)
13. J. Kozák, M. Neˇ, and D. Jan, (2013)
14. J. Debattista, J. Attard, and D. O. Sullivan, 850 (2019)
15. L. L. Idowu and I. I. Ali, 352 (2018)